

A Deep Chatbot for QA and Chitchat

Geonmin Kim^{1*}, Hwaran Lee^{1*}, CheongAn Lee², Eunmi Hong³, Byunggeun Kim¹,
Soo-Young Lee¹

¹School of Electrical Engineering, Korea Advanced Institute of Science and Technology

²AIBrain, ³Crosscert, South Korea

Developing intelligent and natural dialog systems has been a long-term goal of machine learning research. In this work, we build an article-based dialog system that can answer questions about given Wiki articles, as well as, that can engage in casual chitchat with humans. Users subjectively evaluate the system in three metrics: quality, breadth, and the bot’s engagement.

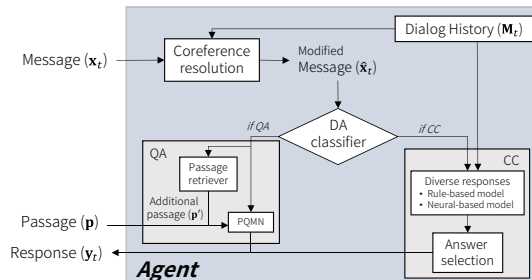


Figure 1. Overall architecture of the system

Table 1. Detail setting of each modules

Module	Model	Database ^a
Coreference Resolution	mention classifier + score matching	cn,co
Dialog Act Classifier	passage match feature + CNN	sq, op, co
Passage Retriever	BM25F [5]	wi
Passage Question Matching Network	Passage RNN + Question RNN + attention-based matching	sq
Chitchat-seq2seq	seq2seq	op,dd, co
Chitchat-rule	ALICE [3] + customized rules	
Answer Selection	Bag-of-words + MLP	op,dd,co

^a cn : CoNLL2012, co : ConvAI, sq : SQuAD [1],

op : Opensubtitle, wi : Wikipedia, dd : DailyDialog [4]

The proposed dialog system consists of four modules as shown in Figure 1: coreference resolution (CR), dialog act classifier (DA), question-answering (QA), and chit-chat (CC). Detail training settings of each module are summarized in Table 1. Firstly, the CR identifies all noun phrases that are referring to an entity, called mentions, and replaces them to the exact entities from the dialog history. It helps for the following modules to understand the user’s message clearly. Then, the DA classifies whether the message requires a question-answering or chit-chat function. According to the classification results, the corresponding module generates a response.

To answer the factual questions about the articles, the QA employs the Passage-Question Matching Network (PQMN) that is trained on SQuAD [1]. The network is matching the questions and passages by attention mechanisms and finds the answer spanning section. Sometimes, users ask factual questions beyond the provided articles. In order for that, we collected another wiki article database and retrieve the related passage by a passage retriever. These additional passages are also fed into the PQMN to answer the question.

To answer the chitchat type messages, the CC firstly generates diverse response candidates by employing both a rule-based model and sequence-to-sequence (Seq2seq) [2] neural models. Then it selects the most appropriate response among the candidates. In detail, for the rule-based model, we utilize the ALICE [3] and additional rules. For the seq2seq models, we trained diverse neural networks with different data and network architecture, and we generate diverse candidates by a diverse beam search technique. We trained an answer selection network which scores each candidates with respect to the appropriateness of the candidate, input message, and dialog history.

We have proposed a new article-dialog system framework incorporating rule-based and deep learning-based components for the Conversational intelligence challenge. We expect the proposed system improves the research area of intelligent and natural dialog systems.

*These authors contributed equally.

References

- [1] Pranav Rajpurkar, Jian Zhang, Konstanin Lopyrev, and Percy Liang. SQuAD: 100,000+ Questions for Machine Comprehension of Text, arXiv 2016.
- [2] Oriol Vinyals, and Quoc Le, A Neural Conversational Model, ICML Deep Learning workshop, 2015
- [3] Richard S. Wallace, The Anatomy of A.L.I.C.E, Parsing the Turing Test, 2009
- [4] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, Shuzi Niu, DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset, ICJNLP, 2017
- [5] Jose R. Perez-Aguera, Javier Arroyo, Jane Greenberg, Joaquin Perez Iglesias, Victor Fresno, "Using BM25F for Semantic Search", Proceedings of the 3rd International Semantic Search Workshop, 2010